Efficient Video-Based Packet Multicast Method for Multimedia Control in Cloud Data Center Networks

Nader F. Mir, Abhishek T. Rao, and Rohit Garg
Department of Electrical Engineering
San Jose State University, California
San Jose, CA, 95195, U.S.A
email: nader.mir@sjsu.edu

***Abstract*— The objective of this paper is to demonstrate the best architectural remedy with the required QoS under running that can handle high end real time streaming traffic while multicasting. Since streaming media accounts for a large portion of the traffic in networks, and its delivery requires continuous service, it is required to analyze the traffic sources and investigate the network performance. IP multicasting technology adds tremendous amount of challenge to a network as streamed media delivered to users must be multiplied in volume. In this paper, we first demonstrate a study of the complexity and feasibility of multicast multimedia networks by using different multicast protocols and video sources. In our proposed method, we then create peer-to-peer (P2P) and data center topologies in order to analyze the performance metrics. The implementation and evaluation of the presented methodology are carried out using OPNET Modeler simulator and the various built-in models. Further, we implement performance tests to compare the efficiency of the presented topologies at various levels.**

Keywords-video streaming; cloud data centers; multicast; multimedia; performance evaluation; video codecs.

I. INTRODUCTION

Over the years, major development in the industry have been involved in the integration of various multimedia applications. Delivery of streaming media (video on demand), e-learning with minimum delay and highest quality has been one of the major challenges in the networking industry. Video service providers, such as Netflix, Hulu are constantly changing the architecture in order to service these needs. These service providers face stiff competition and pressure to deliver the next generation of streaming media to the subscribers. The next generation media can be divided into categories: real-time and non-real time. Examples of real time can be live streaming and video conferencing and non-real time can be e-learning and video on demand [1]. The next generation of streaming media [2] involves a large number of subscribers whose delivery is closer aligned with the latest protocols than with the traditional systems. In such cases, it is required that the service providers upgrade their infrastructure and support them [3].

One of the main challenges in the multimedia industry that motivates us to look into it in this paper is multicasting the video streams. IP Multicast is one of the major techniques that can be used for efficient delivery of streaming multimedia traffic to a large number of subscribers simultaneously. Group membership, unicast and multicast routing protocols are mainly required for multicast communications [4]. *Inter Group Membership Protocol* (IGMP) utilized in our study maintains one of the most commonly used multicast protocols at user facility site. IGMP is used to obtain the multicast information in a network. Unicast routing protocols can be distance vector or link state, the latter being preferred due to the dynamic reaction of these protocols to changes in topology. Multicast routing protocols can be integrated with the unicast routing protocol or can be independent of them. Protocols, such as the *Multicast Open Shortest Path First* (MOSPF), depend on the underlying unicast protocols used, whereas protocols, such as *Protocol Independent Multicast* (PIM), are independent on the type of unicast routing protocols used. A combination of IGMP, MOSPF and PIM in sparse mode or dense mode can be used for successful implementation and efficient delivery of multimedia traffic in networks [4].

Multicast routing enables transmission of data to multiple sources simultaneously. The underlying algorithm involves finding a tree of links connecting to all the routers that contain hosts belonging to a particular multicast group. Multicast packets are then transmitted along the tree path from source to destination (single receiver or group of receivers belonging to a multicast group). In order to achieve the multicast routing tree, several approaches have been adopted. Group-shared tree, source based tree and core based tree are some which are explained here.

a. Group-based tree: In this approach a single routing tree is constructed for all the members in the multicast group
b. Source-based tree: This involves constructing a separate routing tree for each separate member in the multicast group. If multicast routing is carried out using source-based approach, then N separate routing trees are built for each of the N hosts in the group [7].
c. Core-based tree: This is a multicast routing protocol, which builds the routing table using a group-shared tree approach. The tree is built between edge and core routers in a network, which helps in transmitting the multicast packets.

MOSPF and PIM use one of the above mentioned approaches in the transmission of packets. As PIM is the

 ISBN: 978-1-61208-669-9

multicast routing protocol used in the implementation, we discuss the working of PIM.

PIM is a multicast routing protocol that is independent of the underlying unicast routing protocols used [7]. PIM works in two modes- dense mode and sparse mode. In the former mode the multicast group members are located in a dense manner and the latter approach has the multicast group members distributed widely. PIM uses Reverse path forwarding (RPF) technique in dense modes to route the multicast packets. In dense mode, RPF floods packets to all multicast routers that belong to a multicast group whereas in a sparse mode PIM uses a center based method to construct the multicast routing table. PIM routers which work in sparse mode sends messages to a center router called rendezvous point. The router chosen to be rendezvous point transmits the packets using the group based tree model. As seen in Fig. 1, the RP can move from a group based tree model to a source based approach if multiple sources are specified.

The rest of this paper is organized as follows: Section II provides a detail of our architecture and its functionality and Section III presents a performance analysis of the designed architectures. Finally, Section IV concludes the paper.

II. NETWORK ARCHITECTURE

Network architecture has been designed from service provider's and user's perspective. Network service providers are concerned with the available bandwidth and utilization of resources whereas end user's main concern is with the delivery of streaming media with lowest time and maximum efficiency. In order to obtain the various parameters that are required for the best design of multimedia network, two network models were implemented and analyzed.

A. Implemented Peer to Peer (P2P) Network Design

Peer to peer network model is a distributed architecture where the application is transmitted between source and destination through peers. Applications such as music sharing, file sharing use peer-to-peer network model for transmitting the data. A peer-to-peer network was built using the values as shown in Table I. The network architecture shown below represents an organizational division where the admin department is the source of multimedia traffic, which is simultaneously streamed to the remaining departments namely the HR, finance and IT. The topology contains two backbone routers connected back-to-back, a video streaming source is configured and stored in the admin department, where the video frames are encoded with a H.264 codec and generating a frame rate of 15-20 frames per sec. The backbone routers are configured with PIM-DM as the multicast protocol that is responsible to carry multicast packets. The topology diagram can be seen in Fig. 2.

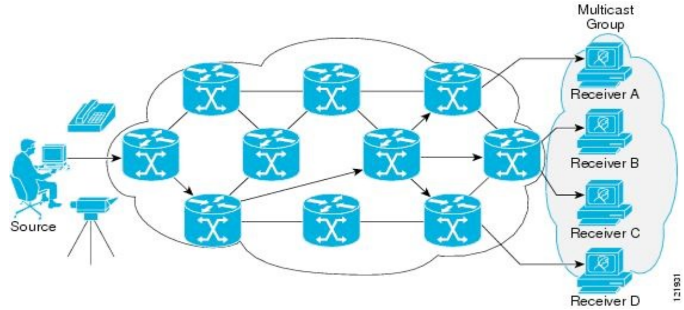

Figure 1. Sample diagram of multicast routing

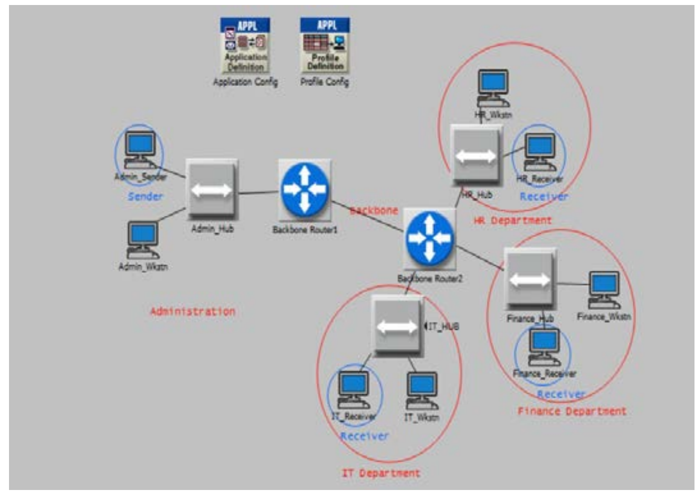

Figure 2. Peer to Peer (P2P) Network Topology

TABLE I. CONFIGURATION PARAMETERS FOR A PEER TO PEER NETWORK DESIGN

Link speed (in Mbps)	Frame size	Frame interarrival rate	Video Codec	Multicast protocol used
100	128x120	10 fps	H.264	PIM-DM
100	128x240	15fps	H.264	PIM-DM
1000	352x240	30 fps	H.264	PIM-DM

B. Implemented Data center topology

A data center contains certain facilities for computing, data storage, and other technology resources distributed over different parts of the world. Data center architecture is divided into access, distribution and core layers. Access layers consist of switches that are connected to servers, distribution layer contains switches which transfer the data from access to core layers, and core layer has high speed switching circuitry that transmits the data over WAN links and to other sites. The data center testbed topology implemented in this paper is shown Fig. 3.

 ISBN: 978-1-61208-669-9

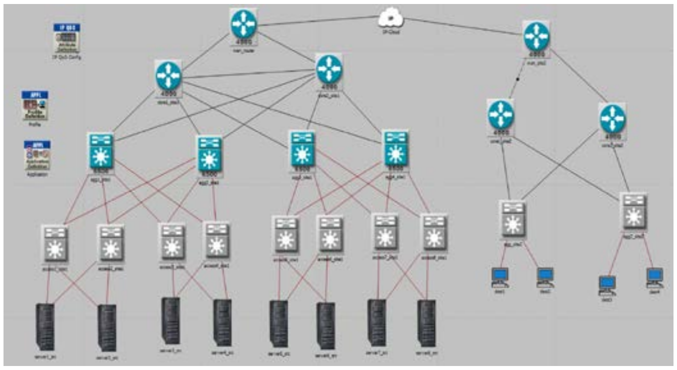

Figure 3. Data Center Topology as the test bed for multimedia streaming

The topology has been implemented taking into account redundancy at all levels. The topology responds dynamically to failures at link, path and device level. Scaling the number of nodes, both horizontally and vertically has been considered in order to analyze the performance metrics of the network. Streaming media content stored at the servers are configured for varying bit rates and varying frame sizes. OPNET simulator has been used as the simulation tool for implementing and testing multicast multimedia traffic detailed metrics that are used for data center implementation has been shown in the Table II.

TABLE II. CONFIGURATION PARAMETERS FOR A DATA CENTER

Number of servers per rack	2
Number of TOR switches used per rack	2
Number of distribution switches per rack	2
Number of core switches per rack	1
Total number of servers	8
Total TOR Switches	8
Total distribution switches	4
Total Core switches	2
Link speeds in data centers	1000 Mbps
Link speeds to WAN	PPP DS3
Video Application and codec used	Video streaming, H.264
Frame sixe	Constant (5000)
Bit rates	Constant (10 fps)

III. PERFORMANCE ANALYSIS

The configuration parameters for used for the performance evaluation are shown below in Table III. Since backbone routers are majorly involved in the transmission of traffic over the internet, Ethernet load across these links has been considered. As the frame size increases load across the backbone links increases, which leads to increase in the delivery of media to destination.

TABLE III. VIDEO CONTENT CONFIGURATION PARAMETERS

Test Name	Frame Size (in bytes)	Video Codec Used	Frame Inter-arrival Time	Ethernet Load Across the Link (packets/sec)
Video1	15,360	H.264	10 Frames/sec	280
Video2	5,000	H.264	Exponential	530

In order to reduce the end to end delay, latency and prioritize traffic Quality of Service (QoS) was implemented. Opnet simulator has various built-in QoS profiles, some of them being WFQ, FIFO, priority queueing. Differentiated services code point based QoS is being used in this implementation wherein based on the priority of traffic delivery, a certain level of service is configured depending on which resources are allocated along the path of delivery.

Now, we present the Ethernet load test – a performance metric which determines the amount of data packets that are carried by the network. Although each link in the network carries data packets WAN / core routers are chosen for analysis. In peer-to-peer topology mentioned earlier, the links connecting the backbone routers are considered, whereas in a data center topology core router links/WAN links have been chosen.

The variation in the graph can be explained as follows. In this case the bit rate and frame size s, both have been kept as exponential increasing functions. From Fig. 4 it can be observed that in a two node network since there is a single link connecting the backbone routers, Ethernet load across these links is considerably higher than that of a multi node model where, PIM builds a tree structure (source based or center based) for sending the multicast packets. As a result, the load is distributed across various links thereby reducing the failure percentage. One more alternative that can be used is port-channel can be configured to distribute the load across the links connecting the routers. Over the time period considered it was observed that the load was higher in a two mode network and lesser in a multi node network.

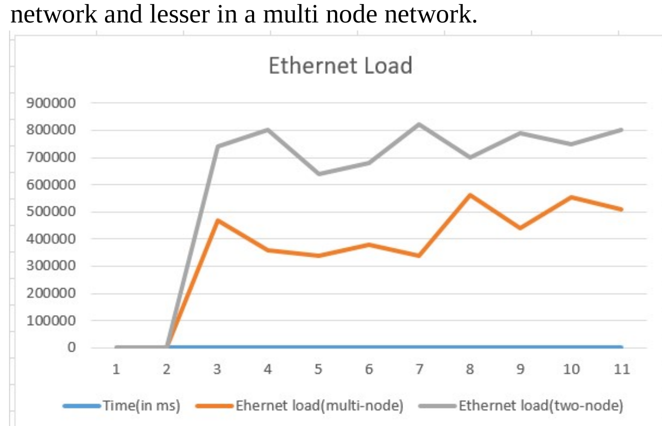

Figure 4. Comparison of Ethernet Load between 2 nodes and multimode case

 ISBN: 978-1-61208-669-9

Our next experiment is concerned with the queueing delay which is the amount of time that a packet waits in the router's queue before being sent onto the network. This is one of the most important parameters for multicast networks as an increase in the queueing delay can cause significant delay in the transmission of packets across the network. Queueing delay can be due to many factors, such as buffer size in a router, router's processing capacity, link speed used, number of hops from source to destination. In this analysis, the queueing delay has been analyzed for a two node and a multi-node environment. From the graphs shown in Fig. 5, it can be observed that although in a two node network links of higher speed were used, when packets of multiple applications arrive, a two node network experienced significant queueing delay which led to the delay in the transmission of packets. Since no QoS was configured all the packets were serviced based on packet arrival times. The graph for a 2 node network shows peaks of highest queueing delay and lows of least queueing delay. This is due to the fact that when packets related to multiple applications arrive there has been peaks of high queue delay and when packets related to single applications arrive less queueing delay has been experienced. In order to have less queueing delay priority traffic can be classified based on QoS policies which helps in serving these packets better.

Next, we consider a test on QoS - a mechanism which is used to analyze the performance of networks. QoS policies configured ensures traffic prioritization and reservation of resources along the path from source to destination. QoS plays a major role in multimedia networks where defining QoS policies defines the traffic priority when real time multimedia traffic and interactive media is involved. Since these types of traffic have rigid delay constraints defining QoS policies for these types can result in prioritizing them when requests for other traffic are in queue.

Simulation results of QoS implementation is shown in Fig. 6. Since real time interactive media could not be created in a simulation environment, two video sources (video1 and video2) were created and video1was configured with a WFQ QoS profile traffic group of video 1 being set to high priority and traffic group of video 2 being set to best effort with no QoS configured. From the plots in the figure, it can be observed that over a period of time when requests arrive for video1 and video2 packets requesting information, video1 is serviced with less packet delay than those packets for video2 while multicast flow is also included in the configuration. Since the IGMP convergence time was 2 min the QoS traffic servicing has started after the first few minutes.

Finally, the latency is our last performance metric to focus on. The latency is the amount of delay involved in transmitting the data from source to destination. For calculating the Latency issues in network different pixel sizes were chosen for analysis. Three different Pixel sizes were configured over a period of time with link speed and other parameters being kept constant. The link speed was defined to be 100 Mbps and pixel sizes of 352x240, 128X240 and 128X120 were defined with frame interarrival rates to be logarithmic. After several tests it was observed that the latency in the transmission of a high quality video was more compared to the latencies of the transmission of a video of lesser resolution as shown in Fig. 7. If a video of high quality has to be transmitted in minimum time, then separate channels can be used for high definition video where source specific trees can be used for routing thereby achieving successful routing of packets.

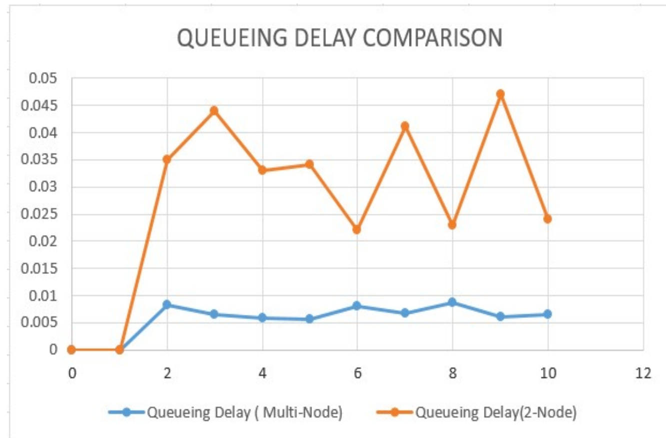

Figure 5. Comparison of queueing delay between 2 nodes and multimode case

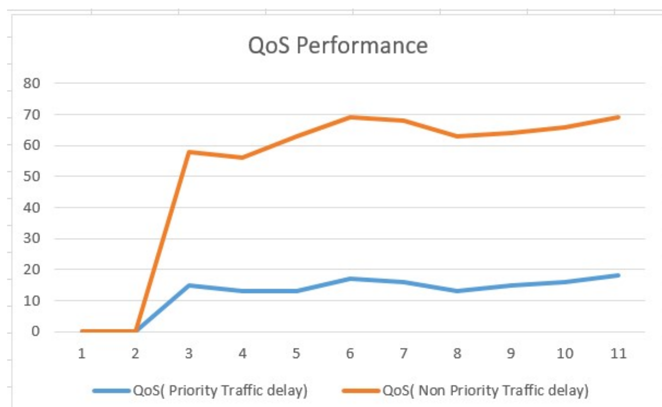

Figure 6. QoS servicing of priority and non-priority traffic

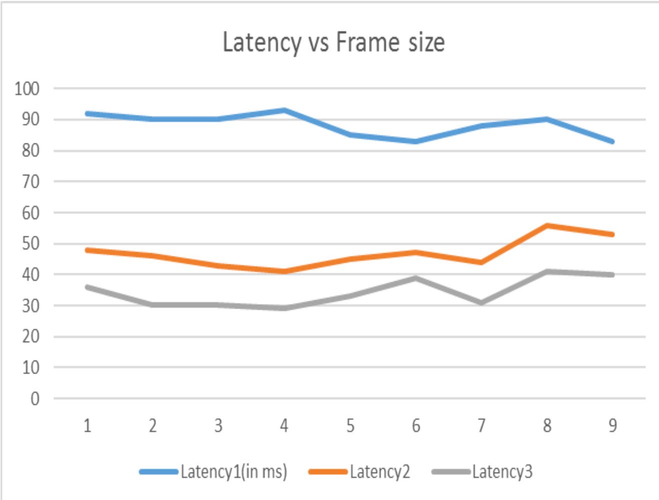

Figure 7. Comparison of Latency of various frame sizes.

IV. CONCLUSION

In this paper, we designed and implemented peer to peer and data center topologies. The two topologies were implemented for various video streaming applications such as video conferencing and video streaming. The parameters of these video sources were changed in to measure the performance metrics of the multicast networks. Parameters such as video codecs, frame size, frame interarrival rate, link speed, QoS were changed for analysis. From the analysis it was observed that a multitier architecture connected to high speed links was best suited for high end real time traffic. Further it was observed that the QoS configuration for these real time traffic reduces the packet end to end delay and the latency of these packets was also less as compared to other packets. Building a multitier not only helped in better load distribution of traffic across links but also this type of topology was better equipped to handle failures at device, links and server levels. This paper also covered the different multicast routing protocols that can be used and how the routing table can be constructed.

REFERENCES

[1] K. Oe, A. Koyama, and L. Batolli, “erformance evaluation of multicast protocols for multimedia traffic,” IEEE Xplore document, IEEE 27th International Conference on Advanced Information Networking and Applications (AINA), 2013.

[2] T. Do, K. Hua, and M. Tantaoui, “P2VoD: Providing fault tolerant video-on-demand streaming in peer-to-peer environment.” Proc. of the IEEE ICC 2004. Paris: IEEE Communications Society, 2004.

[3] B. Cheng, L. Stein, and J. Hetal, “Grid Cast: Improving peer sharing for P2P VOD,” ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 4, 2008.

[4] B. Quinn and K. Almeroth, “IP Multicast Applications: Challenges and Solutions.” RFC, 2001, ACM Digital Library.

[5] “OPNET Technologies Inc. OPNET Modeller Product Documentation Release 14.5,” OPNET Modeller, 2008.

[6] X. Sun, and J. Lu, “The Research in Streaming Media OnDemand Technology based on IP Multicast,” 3rd International Conference on Computer and Electrical Engineering, 2010.

[7] http://www2.ic.uff.br/~michael/kr1999/4-network/4_08-mcast.htm, Access date: July 2018.

[8] Z. Bojkovic, B. Bakmaz, and M. Bakmaz, “Multimedia Traffic in New Generation Networks: Requirements,” Journal of Control and Modeling, 2016.

[9] S. Khanvilkar, F. Bashir, D. Schonfeld, and A. Khokhar, “Multimedia Networks and Communication,” 2015

[10] X. Wang, C. Yu, and H. Schulzrinne, “IP Multicast Simulation in OPNET,” Semantic Scholar, 2015.

.

 ISBN: 978-1-61208-669-9